International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-2, Issue-6, November 2014

A Technique Approaching for Catching User Intention with Textual and Visual Correspondence

Chaitali J. Pawar, Dhanashree V. Patil

***Abstract*—The rapid expansion in web environment and advancement in technology have led us to access and manage enormous images easily in various fields. Present internet image search engines purely faith on keyword based information over images. Keywords provided by user cannot determine content of images correctly. Resultant images comprised of several disordered and uncertain images. To clarify this doubtfulness in keyword-based image search, it is valuable to use visual details of image. System has been designed to overthrown unpredictability of images such that users aim can be determined by one click internet Image search. User preferred a query image in set of images returned by expanded keyword-based search. This technique provides extra clusters generated by applying combination of candidate words and visual features of images by Scale Invariant Feature Transform (SIFT) algorithm. Weight of image is calculated by providing procedure of weight adaption. Results are refined by re-arranging of images by homogeneity calculation. Message Digest (MD5) hash function is used for detecting and removing duplicate images. Resolution of images is considered for further enhancing quality of images. Intention of user is estimated by combining textual and visual resemblance without utilizing extra efforts. An experimental observation determines expressive enrichment in concern with user justice and significancy.**

***Index Terms*— text based search, visual features, adaption weight, cluster.**

I. INTRODUCTION

Along with the amazing popularity of social web sites, enormous amount of web images has appeared in varied content online. Web image search engines like our favorite Google image search mostly depends on surrounding text features of an image. In expectation of certain type of images, user type keywords. Many of returned images are disordered, noisy or irrelevant. Searching of images relevant to a textual query remains a great challenge.

Manuscript received November 23, 2014.
Chaitali J. Pawar, Department of Computer Science and Engineering, Rajarambapu Institute of Technology, Rajaramnagar 415414, India
Dhanashree V. Patil, Department of Computer Science and Engineering, Annasaheb Dange College of Engineering and Technology, Ashta,416301, India

The keyword based image search leads to ambiguous results due to following reasons: 1) Meaning of the query keywords may be distinct than user's expectation. For query apple, result includes red apple, green apple, apple logo etc.
2) Keywords given by user are too short and unable to describe the content of images correctly. Search results are noisy and composed of images having distinct semantic Meanings [1]. 3) Lack of user knowledge on the textual description of target images results into irrelevant images.
4) Expressing the visual content of target images applying keywords correctly is a challenging task for users.

For impressive utilization of human input, achieving the high level user intention is crucial. Content-based image retrieval with relevance feedback requires more users' effort which makes it unsuitable for commercial systems like Google.To overcome uncertainty in text based image search, it is significant to use visual information of image [2]. Keeping in view of this, system has been proposed such that user's intention can be judged by one click internet image search with minimum user labor. Proposed framework needs the user to provide just single click on a query image. Based on text based search, images from a pool are retrieved. Re-ranking is performed depend on their textual and visual correspondence to query image. MD5 hash function is used for detection and removal of duplicate images. By looking for resolution of images, quality of results is further improved. Proposed approach resolves central issue of determining user's intention from single click image search with less user effort. Different industrial applications [3], [4] manage intentional single click image search standard successfully.

II. RELATED WORK

Keyword or text based internet image search undergoes from lot of ambiguity. Numerous text based internet image search methods [5], [6], [7], [8] are restricted due to truth that content of images are not determined properly by query keywords. Depend on visual content for input images, many algorithms produce annotations that results into poor re-ranking [9]. Some of them need supervised training. Determining set of attributes considering hierarchical relationships [10] for highly variant web images is a ultimate task. Discovering unified visual similarity for common images is a challenging task. Most of Pseudo-Relevance

feedback techniques [11], [12] limit user's effort by extending query image with maximum visually similar images. Semantic gap between query image and other visual inconsistent images results into poor performance. Plain visual features and clustering algorithms [13] exhibited huge future of approach by linking keyword and image content on internet image search. Conceptualization of acquiring sample specific visual similarity [14], [15] was investigated in prior work. This demands developing a particular visual similarity for each sample in fixed image pool. So, it is useless for our approach by reason of modifiable image pool for various query keywords. Many techniques in literature [2] subject to deficiency of duplicate images.

Because of distinctiveness in images and features, understanding high level user intention is difficult. So, preserving with this aspect, system has been proposed to overwhelm uncertainty of images so that user's intention can be find out by single click internet image search. Our approach contributes additional image clusters at the time of keyword extension. For furthermore enhancement in performance of image re-ranking, duplicate images are also removed with the help of MD5 hash function. Quality of results is increased by taking into account of resolution of images.

III. METHDOLOGY

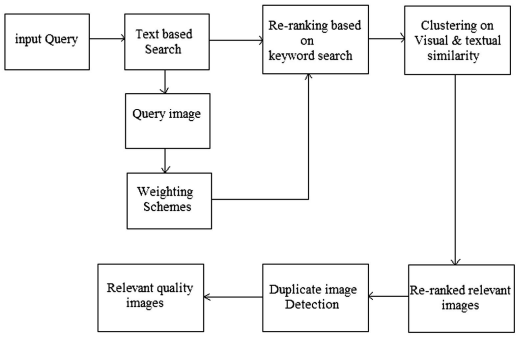

Fig.1. Dataflow of Proposed Approach

Figure1. Shows sequential diagram of our approach. At initial stage, query keyword k is given by user. Depending on the keyword based search, a set of images is returned. From these images, user chooses a query image. Re-ranking is carried out on the collection of images produced over search engine taking account of keyword based search. To determine user intention, keywords are extended rely on visual content of query image preferred by user and over clustering of images. Distinct clusters are produced by considering candidate words and query image descriptor. Re-ranking is performed with help of textual and visual correspondence. Among this collection, duplicate images are eliminated by using MD5 hash function. Results are further enriched in view of resolution of images.

A. Our Method:

Images are manually labeled by user into five categories as a general object, scenery, portrait, object with simple background, or people.

B. Pre-requisite:

Textual information of image (like image_id, image_url, img_keyword, image_description) is valuable in case of keyword extension. Scale Invariant Feature Transform (SIFT) [16] algorithm is beneficial to analyze keypoints (potential interest points) that are invariable to scale and orientation. Rest on the stableness of interest points, keypoints are chosen. They grant a individual feature to discover their absolute match in huge database of features with great probability. Keypoints of an image extracted by SIFT are maintained in a Xml file (Fig. 2).

Fig.2. Working of SIFT

C. Step 1: Text based image retrieval.

Basically, User gives a search query which will be break down into candidate words. (E.g. candidate words of query 'red apple' are {red, apple}. These words further compare with keywords and description of images. Set of compared images shown to user. Term frequency inverse document frequency (Tf-idf) method [17] will be used to remove common words in the search query. These results are further improved by means of keyword extension.

D. Step 2: Candidate word matching & Image Comparison using SIFT

In second step, from set of images by text based search, user select one image in which he/she is interested (i.e. query image). Returned set of images depends on comparison of candidate words with keywords description. This set of images again compare with query image using SIFT. Finally, it gives set of images which contain matching images with query image.

E. Step 3: Creating Clusters of images.

Whenever user select query image, clusters are created by using candidate words and query image descriptor. Considering, candidate word w_i, images consisting of word w_i are found and are bunched into distinct clusters [2]. Now, rely on visual content of query image, all matched images will be put into clusters $\{c_{i,1}, c_{i,2}, \ldots, c_{i,t_i}\}$. These are visually analogous images to query image. E.g. For clustering, if query image is mango and Candidate words: {yellow, mango}, then clusters are {yellow + query image descriptor, mango + query image descriptor}. Image descriptor refers to visual appearance of image (by SIFT).

F. Strategy of adaption weight

Weight of images is accommodated under distinct categories like general object, object with simple background, scenery, portrait and people. Accommodative

resemblance among image i and j is described as [2] by below equation 1:

$$s^q(i,j) = \sum_{m=1}^{p} \alpha_m^q \, s_m(i,j) \quad (1)$$

where $s_m(i,j)$ is resemblance among image i and j over feature m and α_m^q considered as significance of feature m for category Q_q. Weight calculation of query image is done by algorithm which is accommodation of feature weight learning algorithm (Algorithm 1) [2]. To adjustified to our approach, Steps 3 and 4 are distinct from original feature weight algorithm. Because of click on a particular image, steps 5 and 6 in original feature weight algorithm are not required for our approach and described as below:

Algorithm 1: Feature weight learning for a certain query Category.

1. **Input:** Initial weight D_i for all query images i in current intention category Q_q, similarity matrices $s_m(i,.)$ for all query image i and feature m;
2. Initialize: Set step $t = 1$, set $D_i^1 = D_i$ for all i;
While not converged **do**
for each query image $i \in Q_q$ **do**
3. Select best feature m_t (SIFT) and the corresponding Similarity $s_{mt}(i,.)$ for current re-ranking problem under Weight D_i^t;
4. Compute whole weight α_t adjustified with equation 1 in feature weight learning algorithm [2], for our approach, ($\alpha^{t-1} = 0$, *for* $e^{\alpha^{t-1}}$ so, $e^{\alpha^{t-1}} = 1$) by below equation 2

$$\alpha^t = \frac{1}{2} \ln\left(\frac{1+r^t+1}{1+r^t-1}\right) \quad (2)$$

5. t++
end for
end while
6. Output: Final optimal similarity measure for current intention category: $s^q(.,.) - \frac{\sum_t \alpha_t s_{m_t}(.,.)}{\sum_t \alpha_t}$, and weight for feature m: $\alpha_m^q = \frac{\sum_{m_t=m} \alpha_t}{\sum_t \alpha_t}$.

G. Step 4: Re-ranking of images along with visual and textual correspondence.

Re-ranking is carried out on set of similar images with respect to their visual correspondence (no. of keypoints matched by SIFT) with query image. This results into most visually similar images with less irrelevant images.

H. Step 5: Detection of duplicate images.

Resultant set may consist of few duplicate images. It is significant to identify and eliminate duplicate images. Consider, set of visually similar images, S= {a1,a2,a3,a4,a5}, For removing duplicate images, a1 is compared with a2,a3,a4,a5. Again compare a2 with other a3, a4, a5, and so on. Above technique is not suitable for our approach due to more time complexity.

Instead, use of MD5 (message digest) hashing technique is significant. It calculate 128 bit hash value of image. Every pair of nonidentical images will convert into absolutely distinct hash values. Set of images with no duplicate or less duplicate images will be output for this algorithm. Consider, SET_IMG as a set of images and DICT as a dictionary for storing hash of each image and image_url.

Algorithm 2 : Duplicate image Detection.
1. Input: : SET_IMG, DICT **;**
2. For each IMAGE in SET_IMG
3. Calculate MD5 hash of image;
4. Check calculated hash is already in dictionary or not;
5. If hash of image already in DICT it will not added as it is a duplicate image;
end for
6. Output: Set of images with no duplicate images.

This Method is fast and efficient due to less time complexicity.

I. Step 6: Obtaining relevant quality images

Resolution of images are considered for further improvement.

Algorithm 3: Quality Improvement
1. input: Set of images sorted by visual correspondence (no. of keypoints matched by SIFT) with query image; .
2. Maintain resolution of image at retrieval from storage.
3. Set of images will sort depend on count of correspondence of images and highest to lowest resolution of images .
4. Output: Set of images sorted by both image correspondence and quality.

IV. EXPERIMENTAL EVALUATION

For assessment, topmost images are dragged from google image search by applying keyword as query. Distinct 5 classes (like general object, object with simple background, scenery, portrait or people) are formed by manually labeling images (about 500 web images). For estimating user's intention can be determined properly or not, investigation is carried out. Other progressive visual features elaborated in future can be integrated within our approach.
In case of, TABLE I , Figure.4, Figure.5, Let, N=number of images in DB which contains query keyword, **M**= Matching probability of query image and DB image's containing query keywords, **K**= number of keypoints matched in comparing these images. From TABLE I and Figure.4, conclusion will be drawn, when user clicks on the particular query image, result (R) is directly proportional to number of images in DB which contains query keyword. i.e. **R α N.**

From TABLE II and Figure.5, conclusion will be derived; Matching probability of query image and DB image's containing query keywords is directly proportional to number of keypoints matched in comparing these images. Result of query image is also depends on clustering. i.e. **M α K.**

A. Precisions over distinct steps:

To measure performance of re-ranking of image, use of portion of relevant images (topmost m precision) is significant.

Precision (m) = No. of relevant images retrieved by search engine / Total no. of images retrieved. (3)

With the purpose of estimation of effective performance of distinct steps of our approach, comparison of subsequent approaches is carried out. In Figure. 6, average topmost m precisions are considered.

1. TextBased: Google text based search is as a initiation. Term frequency inverse document frequency (Tf-idf) technique is considered for eliminating general words in search query.

2. Visual: Image comparison by SIFT and Candidate word matching are considered for visual correspondence.

3.Visual+text: Image re-ranking is performed in accordance with textual and visual correspondence to query image. Clusters are generated by using candidate words and query image descriptor. For combining visual features, image re-ranking is further enhanced with estimation of weights for images.

TABLE I.

img1	img2	time	Similarities
130120898180264822886616694 1images(3).jpg	100154718237574181563148 05images(8).jpg	11.8336768	5
130120898180264822886616694 1images(3).jpg	889090133675253909475488 52images(15).jpg	10.7546152	10
130120898180264822886616694 1images(3).jpg	208055216816527267752971833 1images(9).jpg	8.5214874	12
130120898180264822886616694 1images(3).jpg	638709000616566274529718331 images(9).jpg	12.9947433	12

TABLE II.

Query image	Results of determining user intention with duplicate image detection & quality improvement
 **fleshy** stone **mango** fruit for **panha** Similar	 **fleshy** stone **mango** fruit for **panha** Similar  indian **mango** with sweet smell Similar  common **mango** for moramba Similar  soft pulpy **mango** for aamras Similar
 **teddy bear** for birthday **gift** Similar	 **teddy bear** for birthday **gift** Similar  **teddy bear** as a soft toy Similar  compliment friends with these cuddly **bears** Similar  **teddy bear** as a iconic children's toy Similar
 a new **born baby girl** with **smile** Similar	 a new **born baby girl** with **smile** Similar  a new **born baby** in summer Similar  a sleeping new **born baby** suffer from pneumonia Similar  premature new **born baby** Similar
 **nature** as a physical **universe** with two **dolphins** Similar	 **nature** as a physical **universe** with two **dolphins** Similar  **nature** referring realm of various types of plants and animals Similar  **nature** pointing to land surface water Similar  dividation of living things into plants and animals by **nature** Similar

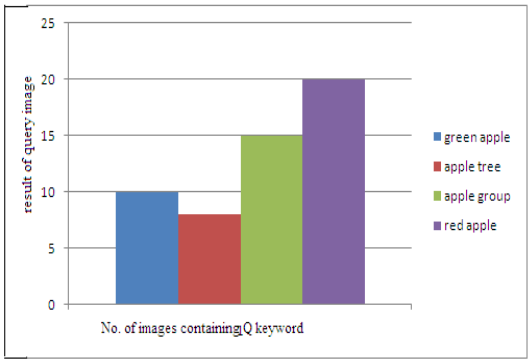

Fig. 4. Comparison of images containing query keyword with it's result .

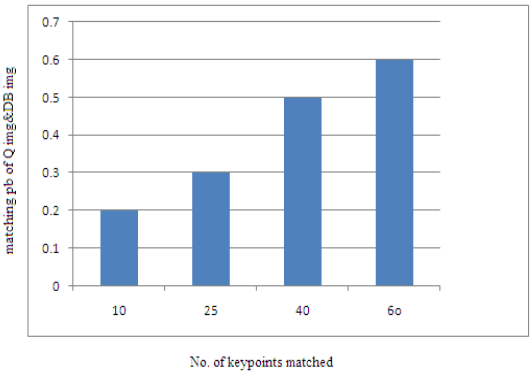

Fig. 5. Comparison of matched keypoints with their probability

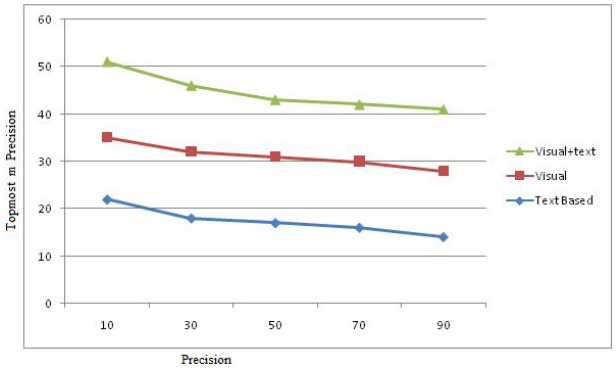

Fig.6. Comparison of average topmost m precisions on distinct steps.

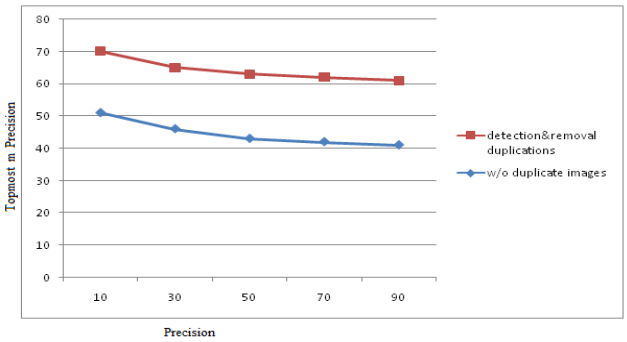

Fig. 7. Comparison of average top m precisions with duplicate image detection and removal.

Fig.6. Clearly shows, TextBased image search results are improved because of utilization of single query image and keyword extension. On account of use of visual features by SIFT, top 50 precision is improved from 17.2 to 31.7 percent. Consideration of textual and visual correspondence (Visual+text) which catches absolute user intention, outperforms Visual property. Topmost 50 precision of Visual is improved from 31.7 to 42.9 percent. At top level, these steps determine user intention effectively. In Figure. 7. Topmost m precisions are considered for duplicate image detection.

1. w/o (detection of) duplicate images: Results of Visual and textual correspondence with query image may consist of duplicate images.

2. Detection and removal of duplicate images: Duplicate images are detected and removed as a consequence of MD5 hashing technique.

Fig.7. purely represents, In concern with detection and removal of duplication, Top 50 precision is improved from 52 to 62.5 percent. This step definitely improves performance.

B. Consultation

In above paper, user intention can be determined properly under presumption that image is both visually and textually correspondence with query image. In some cases, where user is attracted in part of image, intention cannot be properly determined. This part is not considered in above paper due to more user effort.

In accordance with user study and our experimental evaluation, multiple steps can significantly improve performance of re-ranking under condition that top ranked images consist of few relevant images. In some cases, poor initial re-ranking diminishes performance due to improper extension.

V. CONCLUSION

Motivated by idea that utilization of visual details of image is worthy to resolve uncertainty in text-based image retrieval, we have presented a modern Internet image search strategy which necessitate single click of user response. Visual and textual extensions with Tf-idf are incorporated to determine user intention except extra user effort. Evaluation of visual resemblance with query image is accomplished in accordance with image comparison by SIFT and weight strategy. Outcomes of query image are also relying on clustering of images. Re-ranking of images is further enhanced with duplicate image detection by MD5 hashing technique as well as with quality refinement. Our approach gives a chance to commercial scale image search through extended keyword and visual confine. Proposed approach with their subsequent steps can be regenerated by different methods equally in future work. Our strategy can be further refined in future work through application of query log data.

REFERENCES

[1] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W. Ma, "Igroup: Web Image Search Results Clustering," Proc. 14th Ann. ACM Int'l Conf. Multimedia, 2006.

[2] Xiaoou Tang, Ke Liu, Jingyu Cui, Fang Wen and Xiaogang Wang, "IntentSearch : Capturing User Intention for One-Click Internet Image Search", IEEE Transactions on pattern analysis and machine intelligence, vol. 34, no. 7,july 2012.

[3] J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 16th ACM Int'l Conf,Multimedia, 2008.

[4] J. Cui, F. Wen, and X. Tang, "IntentSearch: Interactive On-Line Image Search Re-Ranking," Proc. 16th ACM Int'l Conf.Multimedia,2008.

[5] N. Ben-Haim, B. Babenko, and S. Belongie, "Improving Web Based Image Search via Content Based Clustering," Proc. Int'l Workshop Semantic Learning Applications in Multimedia 2006.

[6] R. Fergus, P. Perona, and A. Zisserman, "A Visual Categor Filter for Google Images," Proc. European Conf. Computer Vision,2004.

[7] Y. Jing and S. Baluja, "Pagerank for Product Image Search," Proc. Int'l Conf. World Wide Web, 2008.

[8] W.H. Hsu, L.S. Kennedy, and S.-F. Chang, "Video Search Re-ranking via Information Bottleneck Principle," Proc. 14th Ann.ACM Int'l Conf. Multimedia, 2006.

[9] C. Wang, F. Jing, L. Zhang, and H. Zhang, "Scalable Search Based Image Annotation of Personal Images," Proc. Eighth ACM Int'l Workshop Multimedia Information Retrieval, 2006.

[10] J. Deng, A.C. Berg, and L. Fei-Fei, "Hierarchical Semantic Indexing for Large Scale Image Retrieval," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2011.

[11] R. Yan, E. Hauptmann, and R. Jin, "Multimedia Search with Pseudo-Relevance Feedback," Proc. Int'l Conf. Image and Video Retrieval, 2003.

[12] R. Yan, A.G. Hauptmann, and R. Jin, "Negative Pseudo- Relevance Feedback in Content-Based Video Retrieval," Proc.11th ACM Int'l Conf. Multimedia, 2003.

[13] B. Luo, X. Wang, and X. Tang, "A World Wide Web Based Image Search Engine Using Text and Image Content Features," Proc. IS&T/SPIE Electronic Imaging, Internet Imaging IV, 2003.

[14] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification," Proc. IEEE Int'l Conf. Computer Vision, 2007.

[15] Y. Lin, T. Liu, and C. Fuh, "Local Ensemble Kernel Learning For Object Category Recognition," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2007.

[16] D. Lowe, "Distinctive Image Features from Scale Invariant Keypoints, " Int'l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[17] R. Baeza-Yates and B. Ribeiro-Neto, "Modern Information Retrieval", Addison-Wesley Longman Publishing Co., 1999.

Prof. Chaitali J. Pawar, has Bachelors and Master's Degree in the field of Computer Science & Engineering,
Area Of Interest : Image Processing, Data Mining, Web Mining, Cloud Computing

Prof. Dhanashree V. Patil, has Bachelors and Master's Degree in the field of Computer Science & Engineering,
Area Of Interest : Image Processing, Cloud Computing Data Mining.